

# Extending FO(ID) with Knowledge Producing Definitions: Preliminary Results

Joost Vennekens<sup>1</sup> and Marc Denecker<sup>2</sup>

<sup>1</sup> Campus De Nayer, Lessius Mechelen, KU Leuven  
De Nayerlaan 5

2860 Sint-Katelijne-Waver, Belgium

<sup>2</sup> Dept. Computerscience, KU leuven  
Celestijnenlaan 200A  
3001 Heverlee, Belgium

**Abstract.** Previous research into the relation between ASP and classical logic has identified at least two different ways in which the former extends the latter. First, ASP program typically contain sets of rules that can be naturally interpreted as inductive definitions, and the language FO(ID) has shown that such inductive definitions can elegantly be added to classical logic in a modular way. Second, there is of course also the well-known epistemic component of ASP, which was mainly emphasized in the early papers on stable model semantics. To investigate whether this kind of knowledge can also, and in a similarly modular way, be added to classical logic, the language of Ordered Epistemic Logic was presented in recent work. However, this logic views the epistemic component as entirely separate from the inductive definition component, thus ignoring any possible interplay between the two. In this paper, we present a language that extends the inductive definition construct found in FO(ID) with an epistemic component, making such interplay possible. The eventual goal of this work is to discover whether it is really appropriate to view the epistemic component and the inductive definition component of ASP as two separate extensions of classical logic, or whether there is also something of importance in the combination of the two.

## 1 Introduction

Today, Answer Set Programming (ASP) is a vibrant domain, boasting both mature technologies and successful real-world applications. The roots of ASP lie in the fields of Logic Programming (LP) and Non-monotonic Reasoning (NMR). Both of these were initially motivated by dissatisfaction with classical first-order logic (FO), be it its computational properties (in the case of LP) or its suitability for representing common-sense knowledge (in the case of NMR). The current success of ASP suggests that, to a large extent, this domain was indeed able to overcome these problems of classical logic.

It is, however, not yet quite clear *how* precisely this was done. That is to say, the relation between ASP and classical logic is, in our opinion, not yet fully

understood. Currently, ASP still stands as an *alternative to* FO: to effectively write ASP programs, one basically has to leave behind all methodologies, tools and intuitive understandings of classical logic and start anew in a different setting. This paper is part of a research project that attempts to close this gap [10]. The aim is to investigate whether and how the achievements of ASP can be reformulated as modular improvements or extensions of classical logic. Ultimately, we would like to be able to characterize ASP as a set of specific solutions to a number of orthogonal problems/limitations of classical logic, such that someone working in classical logic could add as many or as few “ASP-style features” to his knowledge base as is needed for that particular application. Of course, the motivation for this research is not purely practical. By reformulating the contributions of ASP in the classical framework, we also hope to provide a synthesis that will eventually lead to an increased understanding of classical and computational logic, and their role in problem solving.

Ironically, ASP’s relation to classical logic seems currently best understood when it comes to computational aspects. For instance, [20] showed that the typical ASP way of encoding search problems can be captured quite elegantly in a classical context by the notion of *modal expansion*: given a theory  $T$  in an alphabet  $\Sigma$  and an interpretation  $I_o$  for some subvocabulary  $\Sigma_o \subseteq \Sigma$ , find an interpretation  $I$  that extends  $I_o$  to the entire vocabulary such that  $I \models T$ . Indeed, the 2011 edition of the ASP-competition [9] has had two modal expansion systems for (extensions of) classical logic among its competitors: Enfragmo [1] and IDP [19]. On a more technical level, also the similarities between current ASP solvers and SAT solvers are of course well understood (e.g., [16,11,15]).

When it comes to the knowledge representation properties of ASP (i.e., the intuitive meaning of expressions and the modeling methodologies that have to be followed), the relation to classical logic is less clear. As pointed out by [5], one of the key problems lies in the interpretation of the semantic structures: whereas an answer set in ASP is traditionally interpreted in an epistemic way, as a representation of the knowledge of some rational agent, classical logic is based on the Tarskian view of a model as a representation of a possible objective state of affairs. Nevertheless, a series of papers by Denecker et al. [4,8,7] has shown that a substantial portion of ASP can be understood as a combination of classical FO axioms and *inductive definitions*.

An inductive definitions is a well-understood mathematical construct, that is usually represented by a set of natural language *if-then*-statements. As shown in [4,8,7], we can view a set of normal logic programming rules as a formal representation of such an inductive definition. For instance, the following pair of rules:

$$\left\{ \begin{array}{l} T(x, y) \leftarrow T(x, z) \wedge T(z, y). \\ T(x, y) \leftarrow E(x, y). \end{array} \right\}$$

define  $T$  as the transitive closure of  $E$ . This is of course not overly surprising and it indeed goes back to the views of Van Gelder [21] and Clark [2]. Nevertheless, taking this observation seriously immediately suggests a clean and well-defined “plugin” that can modularly add an ASP-style component to a classical FO the-

ory. Indeed, because an inductive definition is nothing more than a generalization of the way in which relations are usually (non-inductively) defined by means of an FO equivalence, there should be nothing problematic (either conceptually or mathematically) about allowing such “sets of rules that form an inductive definition” anywhere one is allowed to use an equivalence. FO(ID) (previously known as ID-logic) is the language that does precisely this: it extends FO, in a completely modular way, with a rule-based representation for inductive definition [3]. Representing a search problem as a model expansion problem in FO(ID) often yields results that are almost identical to a Generate-Define-Test program in ASP [17], apart from minor syntactic details [18,5]. In this way, FO(ID) therefore fits nicely into our stated goal, by identifying one concrete way in which ASP improves upon FO and packing this up in a language construct that can be added to an existing FO knowledge base at will.

While FO(ID) seems able to naturally represent already a surprisingly large part of existing ASP practice, it by no means covers everything. One class of examples that remains out of scope is that of the epistemic examples that originally motivated the stable model semantics [12,13]. This is epitomized by the well-known interview example [14], which is expressed in ASP as:

$$\begin{aligned} Eligible(x) &\leftarrow HighGPA(x). \\ Eligible(x) &\leftarrow Minority(x), FairGPA(x). \\ \neg Eligible(x) &\leftarrow \neg FairGPA(x). \\ Interview(x) &\leftarrow not\ Eligible(x), not\ \neg Eligible(x). \end{aligned}$$

Recent efforts have attempted to reformulate also this kind of example in a way that explains its relation to FO. In [23], the language of *Ordered Epistemic Logic* (OEL) is developed for this. Here, an ordered set of FO theories is considered, and each theory  $T$  is allowed to make use of modal operators  $K_{T'}$  that refer to the knowledge entailed by a theory  $T' < T$ . The interview example would consist of two theories  $T_1, T_2$  with  $T_1 < T_2$ , such that  $T_1$  is a normal FO knowledge base containing facts about *Minority*, *HighGPA* and *FairGPA*, and a definition of *Eligible*. The theory  $T_2$  consists of the single equivalence:

$$\forall x\ Interview(x) \Leftrightarrow \neg K_{T_1} Eligible(x) \wedge \neg K_{T_1} \neg Eligible(x).$$

This logic extends FO in a way which is completely orthogonal to FO(ID). There is nothing to prevent the different knowledge bases of an OEL theory from containing, in addition to regular FO formulas, also inductive definitions, but if they do, there is no interplay with the epistemic operators. In other words, FO(ID) and OEL both isolate one particular non-classical aspect of ASP, and show how it can be modularly added to FO, but they do so independently. This presupposes that there is nothing of importance in ASP’s *combination* of epistemic and “inductive definition” reasoning. However, it does not seem *a priori* obvious that this is the case.

In this paper, we will therefore investigate how an epistemic component can be added to the inductive definition construct of FO(ID) itself. The key idea

here is to allow both a relation *and* an agent's knowledge about this relation to be defined together in a single *knowledge producing definition*, as we will call it. The semantics of such a knowledge producing definition is defined by a constructive process that creates, in parallel, the relations that are being defined and the agent's knowledge about them. In this way, we obtain a language in which, unlike the FO(ID)+OEL approach, interaction between the epistemic and definitional component is possible. The hope is that such a language might shed more light on the epistemic component of ASP and its relation to both classical FO and the inductive definitions of FO(ID).

The work presented in this paper is still at a preliminary stage, but we will attempt to sketch interesting avenues for future research. Because one of our main design goals is to make our approximate knowledge structures integrate seamlessly with the inductive definition construct as it already exists in FO(ID), we will need to spend some time recalling the details of this, before we can develop our extension.

## 2 Preliminaries: the semantics of inductive definitions

Inductive definitions are constructive. In mathematical texts, they are typically represented by a set of *if-then* statements, which may be applied to add new elements to the relation(s) that is (or *are*, in the case of a definition by simultaneous induction) being defined. The formal representation of such a definition in FO(ID) is by a set of rules of the form

$$\forall \mathbf{x} \, P(\mathbf{t}) \leftarrow \phi, \tag{1}$$

where  $P(\mathbf{t})$  is an atom and  $\phi$  an FO formula. For monotone definitions, the relation being defined is simply the least relation that is closed under application of the rules, and it can be constructed by exhaustively applying them. For non-monotone definitions, the relation being defined is no longer the least relation closed under the rules, and there may, in fact, be many minimal relations closed under the rules instead of a single least one. In mathematical texts, such a non-monotone definition is always accompanied by a well-founded order over which the induction is performed. While the characterization as a least set breaks down, the constructive characterization still works: the defined relation can still be constructed by repeated application of the rules, provided that these rules are applied in an order that respects the given well-founded order of induction. For instance, the standard definition of satisfaction in classical logic is a definition over the subformula order, which means that we may only apply a rule that derives that  $I \models \phi$  for some  $\phi$  *after* all rules that could derive  $I \models \psi$  for a subformula  $\psi$  of  $\phi$  have been applied.

Of course, for a correct inductive definition, it is important that the structure of the rules also respects the well-founded order. For instance, in an definition over the subformula order, it makes no sense for a rule to define whether a formula is satisfied in terms of the satisfaction of a *larger* formula. Therefore, the structure of the rules and the well-founded order are not independent. In

fact, the well-founded order is already entirely implicit in the structure of the rules! As shown in [8] and a series of prior papers, the well-founded semantics (WFS) [22] can actually be seen as a mathematical construct to recover the well-founded order from the structure of the rules.

For simplicity, in the technical material of this paper, we will restrict attention to *ground* formulas only.

Since its original definition, a number of alternative ways of defining the WFS have been developed. One of these is to start from the following method of evaluating a formula  $\phi$  in a pair of interpretations  $(I, J)$ :

- For an atom  $P(\mathbf{t})$ ,  $(I, J) \models P(\mathbf{t})$  iff  $I \models P(\mathbf{t})$ ,
- For a formula  $\neg\psi$ ,  $(I, J) \models \neg\psi$  iff  $(J, I) \not\models \psi$ ,
- For a formula  $\psi_1 \vee \psi_2$ ,  $(I, J) \models \psi_1 \vee \psi_2$  iff  $(I, J) \models \psi_1$  or  $(I, J) \models \psi_2$ ,
- For a formula  $\psi_1 \wedge \psi_2$ ,  $(I, J) \models \psi_1 \wedge \psi_2$  iff  $(I, J) \models \psi_1$  and  $(I, J) \models \psi_2$ .

The crux of this definition lies in the case for negation, which switches the roles of  $I$  and  $J$ , thus ensuring that positive occurrences of atoms are evaluated in  $I$  and negative occurrences in  $J$ . We will call a pair  $(I, J)$  for which  $I \leq J$  an *approximating pair*, because it can be seen as an approximation of the set of interpretations  $K$  such that  $I \leq K$  and  $K \leq J$ . Indeed, if  $(I, J) \models \phi$ , according to the above definition, then  $K \models \phi$  for all such  $K$ . Moreover, if  $K \models \phi$  for at least one such  $K$ , then  $(J, I) \models \phi$ . For pairs  $(I, J)$  such that  $I = J$ , the evaluation  $(I, J) \models \phi$  reduces to classical satisfaction  $I \models \phi$ . Pairs for which this is the case are called *total*.

The WFS can then be defined as the unique limit of a sequence of pairs of interpretations  $(I_i, J_i)_{i \geq 0}$ . This sequence starts from the least precise pair of interpretations  $(\perp_\Sigma, \top_\Sigma)$ , where  $\perp_\Sigma$  is the interpretation in which all atoms is false and  $\top_\Sigma$  the interpretation in which all atoms are true. There are then two acceptable ways of going from  $(I_i, J_i)$  to  $(I_{i+1}, J_{i+1})$ :

- Either  $J_{i+1} = J_i$  and  $I_{i+1}$  is the union  $I_i + \{P(\mathbf{t})\}$ , with  $P(\mathbf{t})$  an atom for which there exists a rule of the form  $P(\mathbf{t}) \leftarrow \phi$  with  $(I_i, J_i) \models \phi$ ;
- Or  $I_{i+1} = I_i$  and  $J_{i+1}$  is such that  $I_i \leq J_{i+1} \leq J_i$  and for all atoms  $P(\mathbf{t})$  in the set difference  $J_i - J_{i+1}$  and all rules  $P(\mathbf{t}) \leftarrow \phi$ ,  $(J_{i+1}, I_i) \not\models \phi$

Intuitively, the first of these two cases allows us to derive the head of a rule once it is certain that its body is satisfied (in the sense that  $K \models \phi$  for all  $K$  approximated by  $(I_i, J_i)$ ). The second case allows us to assume that a set of atoms must all be false if this assumption would make us certain that all bodies  $\phi$  of rules with one of these atoms in the head are false (i.e.,  $K \not\models \phi$  for all  $K$  approximated by  $(I_i, J_{i+1})$ ). The set  $J_{i+1} - J_i$  of atoms that are falsified in this operation is known as an *unfounded set*.

A sequence constructed in this way is called an *induction sequence*. The well-founded model (WFM) is now precisely the unique limit  $(V, W)$  to which all such induction sequences converge. If this WFM is total (i.e.,  $V = W$ ), then the definition completely determines the extension of the predicates it defines. Clearly, this is a desirable property for an inductive definition. Therefore, FO(ID) allows only total models.

So far, we have tacitly assumed that all predicates in the vocabulary  $\Sigma$  are defined by the definition. In mathematics, however, this is rarely the case, since most definitions serve to define some relation(s) *in terms of* some other relation(s). This is also possible in FO(ID). For a definition  $\Delta$  (i.e., a set of rules of form (1)), the predicate symbols appearing in the head of at least one of these rules are called the *defined predicates* of  $\Delta$ . The set of all defined predicates is denoted as  $Def(\Delta)$ . The remaining predicates (i.e., those that belong to  $\Sigma - Def(\Delta)$ ) are called *open*, and the set of all such predicates is denoted by  $Op(\Delta)$ . The purpose of a definition is then to characterize the defined predicates  $Def(\Delta)$  in terms of the open predicates  $Op(\Delta)$ . Formally, this is done by parametrizing the construction process by an interpretation for the open predicates: for an interpretation  $O$  of  $Op(\Delta)$ , an induction sequence *given*  $O$  is defined as a sequence of interpretation  $(I_i, J_i)_{i \geq 0}$ , in which all interpretations  $I_i$  and  $J_i$  extend the given interpretation  $O$ . The starting point of this sequence is the pair  $(I_0, J_0)$  such that  $I_0 = O + \perp_{Def(\Delta)}$  and  $J_0 = O + \top_{Def(\Delta)}$ .

An interpretation  $I$  is then called a *model* of a definition  $\Delta$ , denoted  $I \models \Delta$ , if the unique limit of the induction sequences for  $\Delta$  given  $I|_{Op(\Delta)}$  (i.e., the restriction of  $I$  to the open predicates of  $\Delta$ ) is precisely the total pair  $(I, I)$ .

FO(ID) now consists of classical first-order logic FO extended with these definitions. While some versions of this logic allow boolean combinations of classical formulas and inductive definitions, we will, for simplicity, restrict attention in this paper to FO(ID) theories that consist of precisely one FO formula  $\phi$  and one inductive definition  $\Delta$ . For such a theory  $T = \{\phi, \Delta\}$ , we define that  $I$  is a *model* of  $T$ , denoted  $I \models T$ , iff both  $I \models \phi$  (in the classical sense) and  $I \models \Delta$  (as defined above).

### 3 Knowledge producing definitions

The goal of this paper is take the concept of an inductive definition as it exists in FO(ID) and extend it by allowing definitions that not only define the objective extension of their defined predicates, but, at the same time, also define a rational agent's knowledge about these predicates. A *modal literal* is a formula of the form  $K\psi$  where  $\psi$  is an FO formula. By FO(K), we denote the language that extends FO by allowing modal literals to appear anywhere an atom  $P(\mathbf{t})$  may appear. Note that FO(K) therefore does not allow nesting of the operator  $K$ . A knowledge producing definition  $\kappa$  is a set of rules of either the form

$$\forall x K\psi \leftarrow \phi, \quad (2)$$

or

$$\forall x P(\mathbf{t}) \leftarrow \phi. \quad (3)$$

Here,  $K\psi$  is a modal literal and  $P(\mathbf{t})$  an atom. In both cases,  $\phi$  is an FO(K) formula. Again, in our formal treatment, we will always assume that these rules have already been appropriately grounded. The defined predicates of a knowledge producing definition  $\kappa$  are all the predicates  $P$  that appear in the head of a rule

of form (3). All other predicates, including those that appear only in the formula  $\psi$  of a rule of form (2), are open.

For this formalism, the basic semantic structure will consist of a pair of an interpretation  $I$ , representing the real world, and a set of interpretations  $W$ , representing the agent's knowledge about the world. We call such a pair  $(I, W)$  a *knowledge structure* and call it *consistent* if  $I \in W$ . It is obvious how to evaluate a knowledge formula in such a knowledge structure.

**Definition 1.** For a knowledge formula  $\phi$  and a knowledge structure  $\mathcal{S} = (I, W)$ , we define  $\mathcal{S} \models \phi$  as follows:

- For an atom  $P(\mathbf{t})$ ,  $\mathcal{S} \models P(\mathbf{t})$  iff  $I \models P(\mathbf{t})$ ,
- For a modal literal  $K\psi$ ,  $\mathcal{S} \models K\psi$  iff for each  $J$  in  $W$ ,  $(J, W) \models \psi$ ,
- The other cases are defined as usual.

Like the WFS, which construct a single interpretation  $K$  through a series of increasingly precise approximations by pairs of interpretations  $(I_i, J_i)$ , our semantics will construct a single knowledge structure  $(I, W)$  through a series of increasingly precise approximations of it. Part of these approximations will again be a pair of interpretations  $(I_i, J_i)$ , which form an increasingly precise sequence of approximations of the real extension  $K$  of the defined predicates. At each stage of this approximating sequence, we will also keep track of the agent's knowledge about it. Therefore, at each step  $i$ , we will also have a set  $\mathcal{W}_i$  of pairs of interpretations; if  $(I', J') \in \mathcal{W}_i$ , then this means that the agent considers it possible that  $(I', J')$  occurs somewhere in the real approximating sequence. The way in which we will ensure this property, is to, on the one hand, apply the same derivation rules we apply to the real sequence to the pairs that the agent considers possible. On the other hand, when the agent's knowledge increases due to a modal literal in the head of a rule, this will eliminate some of the possibilities (i.e., some of these pairs  $(I', J')$  are removed from  $\mathcal{W}_i$ ), but it will not change any of the possibilities themselves (i.e., the remaining pairs do not change).

The following will serve as our basic semantic structure.

**Definition 2.** An approximate knowledge structure  $\mathcal{A}$  is a pair  $((I, J), \mathcal{W})$  of an approximating pair  $(I, J)$  and a set  $\mathcal{W}$  of approximating pairs  $(I', J')$ .

Evaluating an FO(K) formula in such an approximate knowledge structure is again a matter of switching the approximating pairs when negation is encountered. Formally, we define that, for an approximate knowledge structure  $\mathcal{A} = ((I, J), \mathcal{W})$ ,

- For an atom  $P(\mathbf{t})$ ,  $\mathcal{A} \models P(\mathbf{t})$  iff  $I \models P(\mathbf{t})$ ;
- For a modal literal  $K\psi$ ,  $\mathcal{A} \models K\psi$  iff for each  $(I', J') \in \mathcal{W}$ ,  $((I', J'), \mathcal{W}) \models \psi$ ;
- For a formula  $\neg\phi$ ,  $\mathcal{A} \models \neg\phi$  iff  $((J, I), \bar{\mathcal{W}}) \models \phi$ , where  $\bar{\mathcal{W}} = \{(J', I') \mid (I', J') \in \mathcal{W}\}$ ;
- The other cases are defined as usual.

We now construct an increasingly precise sequence  $(\mathcal{A}_i)_{i \geq 0}$  of approximate knowledge structures. If we project this sequence unto the approximating pair  $(I, J)$  of each approximate knowledge structures  $((I, J), \mathcal{W})$ , the result will essentially be just a regular induction sequence. The sequence again takes as input a knowledge structure  $(O, \mathcal{M})$  for the open predicates, and its starting point is then the approximate knowledge structure

$$\mathcal{A}_0 = ((O + \perp_{Def(\kappa)}, O + \top_{Def(\kappa)}), \{(O' + \perp_{Def(\kappa)}, O' + \top_{Def(\kappa)}) \mid O' \in \mathcal{M}\}).$$

We construct subsequent elements of the sequence by applying one of the following operations.

**Operation 1.**  $\mathcal{A}_{i+1} = ((I_i + \{P(\mathbf{t})\}, J_i), \mathcal{W}_i)$  where  $\mathcal{A}_i = ((I_i, J_i), \mathcal{W}_i)$  such that there is a rule  $r$  of the form  $P(\mathbf{t}) \leftarrow \phi$  with  $((I_i, J_i), \mathcal{W}_i) \models \phi$ .

This operation is just the obvious analogue to the first operation used in building normal induction sequences. The only difference is that the agent's knowledge  $\mathcal{W}_i$  is dragged along as an additional argument, which is used to evaluate occurrences of modal literals in the rule bodies.

**Operation 2.**  $\mathcal{A}_{i+1} = ((I_i, J_i), \mathcal{W}_{i+1})$  where  $\mathcal{A}_i = ((I_i, J_i), \mathcal{W}_i)$  and  $\mathcal{W}_{i+1} = \mathcal{W}_i - \{(I'_i, J'_i)\} + \{(I'_i + \{P(\mathbf{t})\}, J_i)\}$  such that there is a rule  $r$  of the form  $P(\mathbf{t}) \leftarrow \phi$  with  $((I'_i, J'_i), \mathcal{W}_i) \models \phi$ .

This operation is essentially the same as the previous one, with the only difference being that it is now not applied to the approximating pair  $(I_i, J_i)$ , but to one of the approximating pairs  $(I'_i, J'_i)$  in  $\mathcal{W}_i$ . Where the previous two operations are analogous to the production operation of the normal induction sequence, the next two mimic the unfounded set operation.

**Operation 3.**  $\mathcal{A}_{i+1} = ((I_i, J_{i+1}), \mathcal{W}_i)$  where  $\mathcal{A}_i = ((I_i, J_i), \mathcal{W}_i)$  and  $J_{i+1}$  is such that  $I_i \leq J_{i+1} \leq J_i$  and for all atoms  $P(\mathbf{t}) \in J_i - J_{i+1}$  and all rules  $r$  of the form  $P(\mathbf{t}) \leftarrow \phi$ , it holds that  $(J_{i+1}, I_i) \not\models \phi$ .

Again, this operation can either be performed on the approximating pair  $(I_i, J_i)$ , as above, or on one of the pairs in the set  $\mathcal{W}_i$ , as below:

**Operation 4.** Or  $\mathcal{A}_{i+1} = ((I_i, J_i), \mathcal{W}_{i+1})$  where  $\mathcal{A}_i = ((I_i, J_i), \mathcal{W}_i)$  and  $\mathcal{W}_{i+1} = \mathcal{W}_i - \{(I'_i, J'_i)\} + \{(I'_i, J'_{i+1})\}$  and  $J'_{i+1}$  is such that  $I'_i \leq J'_{i+1} \leq J'_i$  and for all atoms  $P(\mathbf{t}) \in J'_i - J'_{i+1}$  and all rules  $r$  of the form  $P(\mathbf{t}) \leftarrow \phi$ , it holds that  $((J'_{i+1}, I'_i), \mathcal{W}_i) \not\models \phi$ .

The final operation takes care of the effect of knowledge producing rules.

**Operation 5.**  $\mathcal{A}_{i+1} = ((I_i, J_i), \mathcal{W}_{i+1})$  where  $\mathcal{A}_i = ((I_i, J_i), \mathcal{W}_i)$  and  $\mathcal{W}_{i+1} = \{(I, J) \in \mathcal{W}_i \mid (J, I) \models \psi\}$  and there exists a rule  $r$  of the form  $K\psi \leftarrow \phi$  such that  $((I_i, J_i), \mathcal{W}_i) \models \phi$ .

Note that the condition for removing a pair  $(I, J)$  from  $\mathcal{W}_i$  is that  $(J, I) \not\models \psi$ , i.e., that no interpretation approximated by  $(I, J)$  still satisfies  $\psi$ .

The semantics of a knowledge producing definition is now defined in terms of sequences that are constructed by these operations.



**Definition 3.** Let  $\kappa$  be a knowledge producing definition. Let  $(O, \mathcal{M})$  be a knowledge structure describing the agent's initial knowledge about the open predicates. A knowledge derivation sequence is a sequence  $(\mathcal{A}_i)_{i \geq 0}$  of approximate knowledge structures, starting from

$$\mathcal{A}_0 = ((O + \perp_{Def(\kappa)}, O + \top_{Def(\kappa)}), \{(O' + \perp_{Def(\kappa)}, O' + \top_{Def(\kappa)}) \mid O' \in \mathcal{M}\}),$$

such that each  $\mathcal{A}_{i+1}$  is obtained from  $\mathcal{A}_i$  by applying one of the five operations defined above and, to prevent the same operation from being applied again and again,  $\mathcal{A}_{i+1} \neq \mathcal{A}_i$ . Such a sequence is called complete if there is no way to extend it further without violating this condition. It is called sound if every operation that is used to construct  $\mathcal{A}_{i+1}$  from  $\mathcal{A}_i$  remains applicable in all  $\mathcal{A}_j$  with  $j > i$ . It is called total if it terminates in an approximate knowledge structure  $((I, J), \mathcal{W})$  such that  $I = J$  and for each  $(I', J') \in \mathcal{W}$  also  $I' = J'$ .

The condition of totality is borrowed from FO(ID), where, as mentioned, it is used to ensure that each inductive definition correctly and completely defines the relations it sets out to define. It is therefore also a natural requirement in our context.

The condition of soundness is meant to avoid situations in which the sequence ends up contradicting itself, as it might for the following example:

$$\{Kp \leftarrow \neg Kp.\}$$

Here, the fact that the agent does not know  $p$  will produce the knowledge that  $p$ . This is not only conceptually problematic, but also creates the practical problem that the order in which operations are applied might have an effect on the final outcome. For instance, the knowledge definition

$$\left\{ \begin{array}{l} Kq \leftarrow \neg Kp. \\ Kp \leftarrow \neg Kq. \end{array} \right\}$$

has both a derivation sequence that starts with the first rule and therefore ends up knowing  $q$  but not knowing  $p$ , and one that starts with the second rule and ends up knowing  $p$  but not  $q$ .

This can only happen with sequences that are not sound, as the following proposition shows.

**Proposition 1.** All complete and sound knowledge derivation sequences that start from the same knowledge structure  $(O, \mathcal{M})$  terminate in the same approximate knowledge structure  $\mathcal{A}$ .

*Proof (sketch).* At any particular point in the derivation sequence, many operation may be applicable. We have to show that it does not matter which of these we choose. First, because the sequence is sound, we know that even if we do not choose to apply an operation now, we will always get a chance to apply it later. Second, the effect of an operation usually does not depend on when it is

executed. The only exception to this is Operation 5, because the set of approximating pairs may of course change throughout the sequence. However, the only changes to this set are that (1) pairs are removed and that (2) pairs become more precise (i.e., that either the first element  $I$  of a pair  $(I, J)$  becomes larger or that  $J$  becomes smaller, so that fewer interpretations  $K$  lie between  $I$  and  $J$ ). Clearly, (1) is not a problem, because if a pair gets removed later on any way, it does not matter if we already remove it now or not. Also (2) is not a problem, because the condition for removing a pair  $(I, J)$ , namely that  $(J, I) \not\models \phi$ , also implies that, for any more precise pair  $(I', J')$  (i.e., such that  $I \leq I'$  and  $J' \leq J$ ), it will be the case that  $(J', I') \not\models \phi$  as well. The only effect of postponing the application of an Operation 5 to a later stage is therefore that we might end up removing *more* pairs than if we had applied it now. However, if we apply the operation now, then at the later stage we will have another Operation 5 available, namely the one that removes precisely those pairs that form the difference. Since this operation will remain applicable and therefore must eventually be applied, the end result will be the same.  $\square$

Moreover, the only way in which it is possible to obtain unsound derivations is by negated modal literals.

**Proposition 2.** *For a knowledge producing definition in which each body of a rule contain only positive occurrences of modal literals  $K\psi$ , each knowledge derivation sequence is sound.*

*Proof (sketch).* The differences between one approximate knowledge structure  $\mathcal{A}_i = ((I_i, J_i), \mathcal{W}_i)$  and an approximate knowledge structure  $\mathcal{A}_j = ((I_j, J_j), \mathcal{W}_j)$  that occurs later in the derivation sequence are:

- $I_i \leq I_j$  and  $J_j \leq J_i$ : this implies that whenever  $((I_i, J_i), \mathcal{W}) \models \phi$  for some  $\mathcal{W}$  and  $\phi$ , also  $((I_j, J_j), \mathcal{W}) \models \phi$
- $\mathcal{W}_j$  consists of pairs  $(I'_j, J'_j)$  for which there exists a corresponding pair  $(I'_i, J'_i) \in \mathcal{W}_i$  such that, again,  $I'_i \leq I'_j$  and  $J'_j \leq J'_i$ , and therefore whenever  $((I'_i, J'_i), \mathcal{W}) \models \phi$  also  $((I'_j, J'_j), \mathcal{W}) \models \phi$
- for some  $(I'_j, J'_j) \in \mathcal{W}_j$ , there may not exist a corresponding pair in  $\mathcal{W}_i$

Putting the second and third point together, it is obvious that  $\mathcal{W}_j$  always knows everything that  $\mathcal{W}_i$  knows, and possibly more. Therefore, only rule bodies containing a negative occurrence of a modal literal may become false in  $\mathcal{W}_j$  after they were true in  $\mathcal{W}_i$ .  $\square$

This result seems to suggest that it might be possible to impose syntactic constraints on a knowledge producing definition to ensure that it only has sound derivations, by limiting the way in which negated modal literals are allowed to appear. However, it is not enough to, e.g., just require that these definitions are stratified, because gaining new knowledge about one predicate may have “side

effects” where also knowledge about another predicate is produced. For instance, consider the following definition which, at first sight, contains no cycles at all:

$$\left\{ \begin{array}{l} q \leftarrow p. \\ Kq \leftarrow r. \\ r \leftarrow \neg Kp. \end{array} \right\}$$

Even though there are no syntactic cycles, once the agent learns  $q$ , this will also produce the knowledge that  $p$  (since  $q$  is only true in worlds where  $p$  also holds). Because of the  $\neg Kp$  in the body of the rule for  $r$ , this will lead to an unsound derivation.

Our approach in this paper will be to ignore the existence of unsound derivations and view the unique limit of the sound derivations as the semantics of a knowledge producing definition—at least, if this unique limit is total. If the limit is not total (given a particular knowledge structure  $\mathcal{S}$  for the open predicates), then, just like in FO(ID), the knowledge producing definition simply has no models (for that particular  $\mathcal{S}$ ). If no sound derivations exists (for that  $\mathcal{S}$ ), then again the knowledge producing definition has no models (for that  $\mathcal{S}$ ).

## 4 Adding knowledge producing definitions to FO

The previous section defined the concept of a knowledge producing definition in isolation. Of course, our goal is to add this construct to FO(K), in the same way as FO(ID) has added inductive definitions to FO. We will again consider only theories of the form  $T = \{\phi, \kappa\}$ , where  $\phi$  is now an FO(K) formula and  $\kappa$  a knowledge producing definition. We define that a consistent knowledge structure  $\mathcal{S} = (I, W)$  is a *weak model* of  $T$  if  $\mathcal{S} \models \phi$  and the approximate knowledge structure  $((I, I), \{(J, J) \mid J \in W\})$  is the unique total limit of each sound derivation sequence that starts from  $(I|_{Op(\kappa)}, \{L|_{Op(\kappa)} \mid L \in W\})$ .

A problem with these weak models is they might contain knowledge that is not warranted. Consider, for instance, the following example:

$$\{q \leftarrow Kp\}.$$

Here, there is no reason at all for knowing  $p$ , yet this knowledge producing definition has a weak model  $(\{p, q\}, \{\{p, q\}\})$ . To avoid such models, we introduce the following concept.

**Definition 4.** Let  $T = \{\phi, \kappa\}$ , where  $\phi$  is an FO(K) formula and  $\kappa$  a knowledge producing definition. A knowledge structure  $\mathcal{S} = (I, W)$  is a *strong model* of  $T$ , denoted  $\mathcal{S} \models T$ , if  $\mathcal{S} \models \phi$  and the approximate knowledge structure  $((I, I), \{(J, J) \mid J \in W\})$  is the unique total limit of each sound derivation sequence that starts from  $(I|_{Op(\kappa)}, O)$  where  $O$  is the set of all  $L|_{Op(\kappa)}$  such that there exists a  $W'$  for which  $(L, W')$  is a weak model of  $T$ .

By always using the set  $O$  as the set of possible worlds for the open predicates, this definition prevents the knowledge producing definition from arbitrarily knowing more about its open predicates than it should.

## 5 Examples

In the logic we have now defined, it is straightforward to represent the *Interview* example.

$$\left\{ \begin{array}{l} Eligible(x) \leftarrow (HighGPA(x) \vee (Minority(x) \wedge FairGPA(x))). \\ Interview(x) \leftarrow \neg K \ Eligible(x) \wedge \neg K \ \neg Eligible(x). \end{array} \right\}$$

$$\forall x \ HighGPA(x) \Leftrightarrow x = Mary.$$

$$\forall x \ FairGPA(x) \Leftrightarrow x = John.$$

$$Minority(Mary).$$

Here, the soundness condition on the derivation sequence means that we first have to apply the first rule of the knowledge producing definition to exhaustion, before starting with the second rule. There are two possible interpretations for the open predicates of the definition that satisfy the FO part of the theory:

$$O_1 = \{HighGPA(Mary), FairGPA(John), Minority(Mary)\},$$

$$O_2 = \{HighGPA(Mary), FairGPA(John), Minority(Mary), Minority(John)\}.$$

Each strong model will therefore have to be produced by a derivation sequence in which  $W_0 = \{O_1, O_2\}$  is used as the agent's knowledge about the open predicates. Let us first consider a sound derivation sequence that interprets the open predicates by the knowledge structure  $\mathcal{S}_1 = (O_1, W_0)$ . The first step is  $\mathcal{A}_0 = (A, \{A, B\})$  with:

$$A = (O_1 + \perp_{Def(\kappa)}, O_1 + \top_{Def(\kappa)})$$

$$B = (O_2 + \perp_{Def(\kappa)}, O_2 + \top_{Def(\kappa)})$$

In the approximating pair  $A$ , we can apply the first rule of the knowledge producing definition to derive that  $Eligible(Mary)$ , and we can also apply an unfounded set operation to derive that  $Eligible(John)$  is false. This leads to the new approximating pair (we abbreviate the names of the predicates):

$$A' = (O_1 + \{Elig(Mary)\}, O_1 + \{Elig(Mary), Int(Mary), Int(John)\})$$

In the approximating pair  $B$ , on the other hand, we can derive that both  $Eligible(John)$  and  $Eligible(Mary)$ . This leads to the new pair:

$$B' = (O_2 + \{Elig(Mary), Elig(John)\},$$

$$O_2 + \{Elig(Mary), Elig(John), Int(Mary), Int(John)\})$$

After applying these six operations (two times two for  $A$ , and two for  $B$ ) to the approximate knowledge structure  $\mathcal{W}_0$ , we will therefore eventually end up in

$$\mathcal{W}_6 = (A', \{A', B'\})$$

Now,  $\{A', B'\} \models K \text{Elig}(\text{Mary})$ , because  $\text{Elig}(\text{Mary})$  holds in both underestimates  $O_1 + \{\text{Elig}(\text{Mary})\}$  and  $O_2 + \{\text{Elig}(\text{Mary}), \text{Elig}(\text{John})\}$ . Also,  $\{A', B'\} \models \neg K \text{Elig}(\text{John})$ , since  $\text{Elig}(\text{John})$  does not hold in the overestimate (the negation switches the pairs) of  $A'$ , namely  $O_1 + \{\text{Elig}(\text{Mary}), \text{Int}(\text{Mary}), \text{Int}(\text{John})\}$ . Finally,  $\{A', B'\} \models \neg K \neg \text{Elig}(\text{John})$  holds as well, since  $\text{Elig}(\text{John})$  does not belong to the underestimate (the two negations switch the pairs twice) of  $A'$ , namely  $O_1 + \{\text{Elig}(\text{Mary})\}$ . Therefore, we can apply to both  $A'$  and  $B'$  an unfounded set operation (Operations 3 and 4) to derive that  $\text{Mary}$  should not be interviewed, and we can apply the last rule of the definition to derive that  $\text{John}$  should. After six more steps (two times two for  $A'$ , and two for  $B'$ ), we therefore end up in the limit  $\mathcal{W}_9 = (A'', \{A'', B''\})$ , where

$$\begin{aligned} A'' &= (I, I) \text{ with } I = O_1 + \{\text{Elig}(\text{Mary}), \text{Int}(\text{John})\} \\ B'' &= (J, J) \text{ with } J = O_1 + \{\text{Elig}(\text{Mary}), \text{Elig}(\text{John}), \text{Int}(\text{John})\} \end{aligned}$$

This structure is total and therefore the single knowledge structure  $(I, \{I, J\})$  that it approximates is a model of this theory. By a similar reasoning, there is also a second model, namely  $(J, \{I, J\})$ .

### 5.1 Another version of the *Interview* example

In the above version of this example, we are basically already encoding the solution to the problem by ordering the agent to interview everyone whose eligibility is not known. Using a knowledge producing definition, however, it is also possible to let the semantics do more of the work.

$$\left\{ \begin{array}{l} \text{Eligible}(x) \leftarrow (\text{HighGPA}(x) \vee (\text{Minority}(x) \wedge \text{FairGPA}(x))). \\ K \text{Minority}(x) \leftarrow \text{Interview}(x) \wedge \text{Minority}(x). \\ K \neg \text{Minority}(x) \leftarrow \text{Interview}(x) \wedge \neg \text{Minority}(x). \end{array} \right\}$$

$$\begin{aligned} (\forall x \text{HighGPA}(x) \Leftrightarrow x = \text{Mary}) \wedge (\forall x \text{FairGPA}(x) \Leftrightarrow x = \text{John}) \wedge \text{Minority}(\text{Mary}). \\ \forall x K \text{Eligible}(x) \vee K \neg \text{Eligible}(x). \end{aligned}$$

Here, we are just telling the agent that the action of interviewing a candidate will reveal his minority status, without explicitly saying who should be interviewed. The FO(K) constraint then orders the agent to make sure that for each candidate, it is known whether he is eligible or not. Generating models for this theory will then correctly produce plans in which people whose minority status is unknown will be interviewed.

### 5.2 Sensing actions

A successful application area of ASP is planning. This also falls naturally in the scope of FO(ID), since theories in the situation or event calculus are essentially just an inductive definition of the values of the fluents at different points in time

[6]. Here is an example of a theory in FO(ID) that represents a small action domain in which there is a dirty glass that can be cleaned by wiping it.

$$\left\{ \begin{array}{l} \text{Clean}(t+1) \leftarrow \text{Wipe}(t) \vee \text{Clean}(t). \\ \text{Clean}(0) \leftarrow \text{InitClean}. \end{array} \right\}$$

$$\neg \text{InitClean}$$

If the agent now does not know whether the glass is initially clean, we may be interested in finding a plan that will allow it to know with certainty that it will be clean at a certain point in time. This can be accomplished by just adding, e.g.,  $K \text{ Clean}(2)$  as an FO(K) constraint to the theory. More interestingly, knowledge producing definitions can also be used to add *sensing* actions, such as an action *Inspect* that allows the agent to discover whether the glass is clean.

$$\left\{ \begin{array}{l} \text{Clean}(t+1) \leftarrow \text{Wipe}(t) \vee \text{Clean}(t). \\ \text{Clean}(0) \leftarrow \text{InitClean}. \\ K \text{ Clean}(t+1) \leftarrow \text{Inspect} \wedge \text{Clean}(t). \\ K \neg \text{Clean}(t+1) \leftarrow \text{Inspect} \wedge \neg \text{Clean}(t). \end{array} \right\}$$

$$K \text{ Clean}(2).$$

## 6 Discussion

ASP is able to express epistemic examples by interpreting an answer set as a set of literals that are *believed* by a rational agent. This is one of the most radical ways in which ASP departs from classical logic, in which models or interpretations always represent the *objective* state of the world. In the classical setting, one typically resorts to sets of interpretations (or the related concept of a Kripke structure) to represent beliefs.

In order to relate this aspect of ASP to classical logic, or to even integrate the two, it is necessary to construct a formalisation which sticks to these classical semantics objects. In this paper, we have introduced knowledge producing definitions for this purpose. As our examples have shown, these are able to mimic the ASP representation of, e.g., the *Interview* example, while at the same time also introducing some interesting new possibilities, such as the ability to distinguish between some atom  $P(t)$  becoming objectively true (by having  $P(t)$  in the head of a rule) and the agent learning this atom (by having  $K P(t)$  in the head).

## References

1. A. Aavani, X. Wu, E. Ternovska, and D. Mitchell. Enfragmo: A system for modelling and solving search problems with logic. In *LPAR-18*, 2012.

2. K. Clark. Negation as failure. In *Logic and Data Bases*, pages 293–322. Plenum Press, 1978.
3. M. Denecker. Extending classical logic with inductive definitions. In *CL*, volume 1861 of *LNCS*, pages 703–717, 2000.
4. M. Denecker, M. Bruynooghe, and V. Marek. Logic programming revisited: Logic programs as inductive definitions. *ACM Transactions on Computational Logic (TOCL)*, 2(4):623–654, 2001.
5. M. Denecker, Y. Lierler, M. Truszczyński, and J. Vennekens. A Tarskian informal semantics for ASP. In *Technical Communications of the 28th International Conference on Logic Programming*, 2012.
6. M. Denecker and E. Ternovska. Inductive situation calculus. *Artificial Intelligence*, 171(5-6):332–360, 2007.
7. M. Denecker and E. Ternovska. A logic of nonmonotone inductive definitions. *ACM Transactions on Computational Logic (TOCL)*, 9(2):Article 14, 2008.
8. M. Denecker and J. Vennekens. Well-founded semantics and the algebraic theory of non-monotone inductive definitions. In *LPNMR*, pages 84–96, 2007.
9. M. Denecker, J. Vennekens, S. Bond, M. Gebser, and M. Truszczyński. The second answer set programming system competition. In *LPNMR*, pages 637–654, 2009.
10. M. Denecker, J. Vennekens, H. Vlaeminck, J. Wittocx, and M. Bruynooghe. Answer set programming’s contributions to classical logic. An analysis of ASP methodology. In *MG-65: Symposium on Constructive Mathematics in Computer Science*. Lexington, October 26-27, 2010.
11. M. Gebser, T. Schaub. Tableau Calculi for Answer Set Programming. In *ICLP*, pages 11-25, 2006.
12. M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *ICLP/SLP*, pages 1070–1080. MIT Press, 1988.
13. M. Gelfond and V. Lifschitz. Logic programs with classical negation. In *ICLP*, pages 579–597. MIT Press, 1990.
14. M. Gelfond and V. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3/4):365–386, 1991.
15. E. Giunchiglia, N. Leone, M. Maratea. On the relation among answer set solvers. In *Ann. Math. Artif. Intell.* 53(1-4), pages 169-204 (2008).
16. Y. Lierler. Abstract answer set solvers with backjumping and learning. *Theory and Practice of Logic Programming (TPLP)* 11(2-3): 135-169, 2011.
17. V. Lifschitz. Answer set programming and plan generation. *Artificial Intelligence*, 138:39–54, 2002.
18. M. Mariën, D. Gilis, and M. Denecker. On the relation between ID-Logic and Answer Set Programming. In *JELIA*, volume 3229 of *LNCS*, pages 108–120. Springer, 2004.
19. M. Mariën, J. Wittocx, and M. Denecker. The IDP framework for declarative problem solving. In *Search and Logic: Answer Set Programming and SAT*, pages 19–34, 2006.
20. D. Mitchell and E. Ternovska. A framework for representing and solving NP search problems. In *AAAI*, pages 430–435, 2005.
21. A. Van Gelder. The alternating fixpoint of logic programs with negation. *Journal of Computer and System Sciences*, 47(1):185–221, 1993.
22. A. Van Gelder, K. Ross, and J. Schlipf. The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650, 1991.
23. H. Vlaeminck, J. Vennekens, M. Bruynooghe, and M. Denecker. Ordered epis-temic logic. In *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning*, 2012.